Abdominal Radiology (2019) 44:2172–2181
https://doi.org/10.1007/s00261-019-01960-1

HEPATOBILIARY

Comparison of navigator-gated and breath-held image acquisition techniques for multi-echo quantitative dixon imaging of the liver in children and young adults

Leah A. Gilligan[1] · Jonathan R. Dillman[1,2] · Jean A. Tkach[1] · Andrew T. Trout[1,2]

Published online: 27 February 2019

Abstract

Purpose Acquired over a breath hold, multi-echo Dixon (mDixon) magnetic resonance imaging (MRI) of the liver can be used to quantify proton density fat fraction (PDFF) and iron-related signal decay. However, young, obese, and co-morbid patients may have limited breath holding capacity and could benefit from a motion-robust mDixon acquisition. The purpose of this study was to compare hepatic PDFF and R2* values between navigator-gated and breath-held mDixon MRI acquisition techniques in children and young adults with suspected liver disease.

Materials and methods This retrospective study was institutional review board-approved with a waiver of informed consent. Patients who underwent liver MRI with breath-held and navigator-gated mDixon sequences between January 2017 and July 2018 were included. One reviewer, blinded to sequence, measured PDFF and R2* on four images from each sequence. Another blinded reviewer graded respiratory motion (5-point Likert scale). Pearson correlation (r), Lin's concordance coefficients (r_c), and Bland–Altman analyses were used to assess agreement between techniques. Frequency of clinically limiting motion (score ≥ 3) was compared with Fisher's exact test.

Results Forty-two patients were included (15 female, 27 male; mean age: 15.7 ± 4.6 years). Mean PDFF and R2* were $16.6 \pm 13.1\%$ and $29.3 \pm 4.7\ s^{-1}$ (breath-held) versus $17.0 \pm 13.2\%$ and $29.6 \pm 5.2\ s^{-1}$ (navigator-gated). PDFF agreed almost perfectly between sequences ($r_c = 0.997$, 95% CI 0.994–0.998; mean bias: 0.3%; 95% limits of agreement: -2.4 to $+1.7\%$), while R2* values correlated very strongly but with poor agreement ($r = 0.837$, $r_c = 0.832$, 95% CI 0.716–0.910). Navigator-gated images exhibited significantly higher frequency of clinically limiting respiratory motion (88% vs. 48%, $p = 0.0001$).

Conclusion Despite greater respiratory motion artifact, a free-breathing navigator-gated mDixon sequence produces PDFF values with almost perfect agreement to a breath-held sequence.

Keywords Proton density fat fraction · Liver MRI · Respiratory triggering · Pediatric · R2* · T2*

Introduction

Magnetic resonance imaging (MRI) sequences employing a multi-echo Dixon (mDixon) method exploit differences in chemical shift to separate the MRI signals from water and fat. Those signals can then be recombined using image math to generate water-only, fat-only, and in- and opposed-phase images. Carefully crafted mDixon sequences also allow quantification of liver proton density fat fraction (PDFF) and T2* (or R2* = 1000/T2*) relaxation rate, which are increasingly being used to characterize diffuse liver disease in children [1–14]. PDFF values, generated from an mDixon sequence, have been validated as an imaging biomarker of hepatic fat content through correlation with historical reference standards, including steatosis grade determined by histology and fat fraction obtained by magnetic resonance (MR) spectroscopy, and are utilized clinically for the diagnosis and surveillance of non-alcoholic fatty liver disease (NAFLD) [15–18]. Additionally, T2* or R2* values, derived from an mDixon sequence, allow calculation of liver iron

✉ Andrew T. Trout
andrew.trout@cchmc.org

[1] Department of Radiology, MLC 5031, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA

[2] Department of Radiology, University of Cincinnati College of Medicine, Cincinnati, OH, USA

content with results shown to correlate strongly with traditional relaxometry methods [19–21].

Image acquisition of an mDixon sequence typically requires a breath hold to minimize respiratory motion artifact. However, breath holding capacity can be limited and unreliable in children, who generally have higher and more irregular respiratory rates than adults, and in patients of any age with dyspnea or tachypnea [22, 23]. For this reason, motion-robust imaging acquisition techniques, including respiratory compensation with gating using a navigator echo (navigator-gating), would be of value for MR imaging of children and certain adults.

The purpose of this study was to compare quantitative mDixon sequences acquired with breath holding (standard of care) and with navigator-gating in children and young adults with suspected liver disease. Specifically, we sought to: (1) define agreement between measured liver PDFF and R2* values, (2) compare measureable region of interest (ROI) size, and (3) compare severity of motion between the two sequences. We hypothesized that liver PDFF and R2* measured on the navigator-gated images would strongly correlate with the breath-held images and that the navigator-gated acquired images would have less respiratory motion.

Methods

The institutional review board at Cincinnati Children's Hospital Medical Center approved this retrospective, Health Insurance Portability and Accountability Act (HIPAA)-compliant study. A waiver of documentation of informed consent was granted.

All consecutive patients who underwent liver MRI examination for any clinical indication that included acquisition of both a breath-held and navigator-gated mDixon sequence for fat and R2* quantification at our institution between January 1, 2017 and July 25, 2018 were included in the study. Participants were identified through a query of our institution's picture archiving and communication system (PACS) using open-source software (SQuirreL SQL, http://squirrel-sql.sourceforge.net). No patient-level exclusion criteria were applied.

Patient demographic data, relevant medical history, including etiology of liver disease, and clinical imaging report data were derived from a query of the institutional medical record (Epic Systems Corp., Verona, WI). Subject height, weight, body mass index (BMI), and BMI percentile obtained closest to, and within 100 days of, the MRI exam were recorded.

Imaging protocol

All imaging examinations had been performed on HDx 1.5 Tesla (T) MR imagers (GE Healthcare, Waukesha, WI). Detailed parameters for the mDixon (IDEALIQ®, GE Healthcare, Waukesha, WI) sequences (breath-held and navigator-gated) are provided in Table 1. For the breath-held acquisition, patients were instructed to hold their breath in end-expiration (duration of breath hold ~ 25–28 s). For the navigator-gated acquisition, (acquisition time ~ 90–125 s) a navigator echo was implemented as a 20 ms 2D pencil-beam radiofrequency excitation pulse, with its long axis (coincident with navigator read out direction) placed along the superior–inferior (S/I) direction. The navigator RF pulse was applied every 200 ms positioned to fall at the beginning and end of each imaging block. Images were used to guide the placement of the navigator on the most superior aspect of the dome of the liver on the right side, centered anterior to posterior. The navigator was placed so that the S/I extent of the beam was split roughly 1/2–1/2 to 1/3–2/3 between the lung and the liver. The navigator beam was typically 10-15 cm in length and 2 mm in width, with the user defined acceptance window set to ~ 2 mm. Only *k*-space data that was collected during periods when the diaphragm position,

Table 1 MR pulse sequence parameters for the acquired mDixon sequences

Parameter	
Imaging platform	GE HD23
Coil	8—Channel Cardiac
Sequence	IDEAL IQ®
Field strength (*T*)	1.5
Technique/pulse sequence	3D Fast RF spoiled gradient echo
Scan Plane	Axial
Frequency direction	Right/left
FOV (mm)	320–460
Matrix	220–256 × 160
No. of signals acquired	0.724359–0.740909
TE (ms)	Min Full (5.47–7.41)
No. of echoes	6
TR (ms)	11.24–21.525
Flip angle (degrees)	5
Bandwidth (Hz)	125 K
No. of sections	20–40
Section (3D partition) thickness (mm)	8–10
Acceleration	ARC
Acceleration factor	2 phase; 2 slice

The only difference between the sequences was whether they were acquired with a single breath hold or with navigator-gating

FOV field of view; *TE* time to echo, *TR* repetition time

as determined by the navigator, fell within the user-defined acceptance range during end-expiration, were retained and used for image creation.

Image processing—hepatic PDFF and R2*

All PDFF and R2* measurements were performed using a vendor-neutral post-processing platform (IntelliSpace, Philips Healthcare, Best, the Netherlands). For the purpose of this study, PDFF and R2* were measured by a single reviewer, a medical doctor, (LAG) who first performed measurements on all breath-held image series and, 1 week later and blinded to the initial measurements, on all navigator-gated image series. The single reviewer measured hepatic PDFF and R2* by drawing four free-hand regions of interest (ROIs) on each the PDFF map and R2* map images, respectively. ROIs were drawn to include as much liver parenchyma as possible while excluding the liver capsule, biliary tree, large vessels, and discrete areas of artifact (Fig. 1) [24]. For each breath-held and navigator-gated series, four ROIs were drawn, with each ROI placed on a separate axial image spaced throughout the craniocaudal dimension to maximize liver parenchymal sampling. No attempt was made to replicate ROIs across slices within a series or between series. For the purpose of further blinding, recording of PDFF and R2* values from individual ROIs was deferred until all four ROIs were placed. PDFF values, R2* values, and ROI areas were recorded. To ensure accuracy of ROI placement, approximately 20% of the ROIs drawn for study purposes ($n = 136/672$) were reviewed by a board-certified pediatric radiologist with more than 5 years of post-fellowship experience (ATT).

Calculation of hepatic PDFF and R2*

To account for variability in ROI size, PDFF and R2* were expressed as a weighted mean of the four measured mean values, each weighted by its respective ROI size. As a secondary aim, we sought to define the impact of number of ROIs on measured PDFF and R2*, and therefore, also

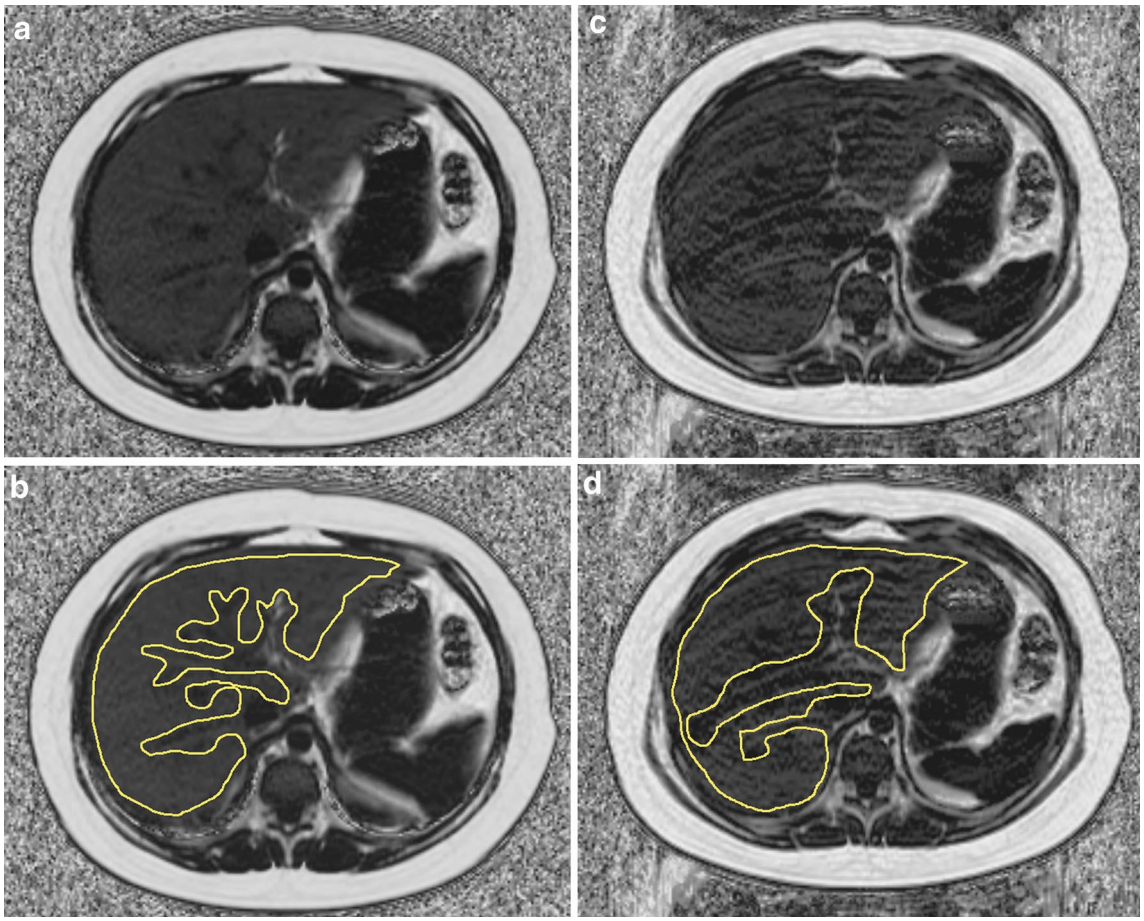


Fig. 1 Selected fat fraction parametric map images of an 11-year-old male from the breath-held sequence (**a**, **b**) and navigator-gated sequence (**c**, **d**) without **a**, **c** and with **b**, **d** ROIs applied. Note the navigator-gated image **c** exhibits clinically limiting motion (score of 4) while the breath-held image **a** does not (score of 2)

calculated PDFF and R2* based on a single ROI and as a weighted mean of two and three ROIs. When PDFF and R2* were calculated using fewer than four ROIs, the first one, two, and/or three ROIs generated for each patient were used.

Imaging analysis—motion

De-identified fat fraction maps were reviewed by one board-certified pediatric radiologist (ATT) who was blinded to image acquisition technique (breath-held versus navigator-gated). The reviewer assigned a respiratory motion-related artifact score using the following scale: score 1, none; score 2, minimal with no effect on diagnostic quality; score 3, moderate with some but no severe effect on diagnostic quality; score 4, severe but images still interpretable; and 5, extensive and images non-diagnostic. For statistical analysis, qualitative motion scoring was dichotomized into "non-clinically limiting" and "clinically limiting" motion, the latter defined as a score of ≥ 3 (Fig. 1).

Statistical analysis

Continuous data were summarized as means and standard deviations; categorical data were summarized as counts and percentages. Student *t* tests (paired and/or unpaired, two-sided) and Fisher's exact test were used for comparisons of means and frequencies, including mean ROI area for PDFF measurement, mean PDFF values, mean R2* values, and frequency of clinically limiting motion between techniques. Pearson correlation (r) and Lin's concordance correlation coefficient (r_c) were used to assess agreement between continuous measures of PDFF and R2* including breath-held versus navigator-gated technique using all four ROIs and using one, two, or three ROI(s) only. Bland–Altman difference plots were created to further assess agreement/bias in measurements of PDFF and R2* between breath-held and navigator-gated techniques. Given the known greater coefficient of variation in measured PDFF at lower PDFF values, analyses of PDFF data were repeated for the subset of patients with a four ROI weighted mean PDFF less than 9% by at least one of the two imaging techniques [25]. The 9% threshold was selected based on visual inspection of the study population data spread for PDFF.

Univariable and multivariable regression were performed to define patient characteristics (age, sex, BMI, PDFF) associated with qualitative clinically-limiting motion. *P* values < 0.05 were considered statistically significant for inference testing; 95% confidence intervals (CIs) were calculated as appropriate. Pearson correlation coefficients were classified by the following definitions: 0–0.19, very weak; 0.2–0.39, weak; 0.40–0.59, moderate; 0.60–0.79, strong; and 0.80–1.0, very strong [26]. Lin's concordance correlation coefficients were classified by the following definitions: less than 0.90, poor agreement; 0.90–0.95, moderate agreement; greater than 0.95–0.99, substantial agreement; and greater than 0.99, almost perfect agreement [27]. All statistical analyses were performed using SPSS (IBM, Armonk, NY), SAS (SAS Institute, Inc., Cary, NC), MedCalc (MedCalc, Ostend, Belgium), or GraphPad QuickCalcs (GraphPad Software, La Jolla, CA).

Results

Forty-two patients had both breath-held and navigator-gated acquisitions of the mDixon sequence performed during the study period and were included in this study. Mean patient age was 15.7 ± 4.6 years. Patient demographic information, four ROI-weighted mean PDFF, and four ROI-weighted mean R2* are summarized in Table 2.

Number and size of ROIs for PDFF

There were very strong-positive linear correlations between PDFF measurements calculated as the weighted mean of one, two, and three ROIs compared to four ROIs for each of the breath-held and the navigator-gated sequences with almost perfect corresponding Lin's concordance correlation coefficients (Table 3, Fig. 2).

Table 2 Demographic characteristics of study population

Parameter	$n = 42$
No. of male patients	27 (64.3)
Age (years)	15.7 ± 4.6
Height (cm)	162.3 ± 17.8
Weight (kg)	86.4 ± 33.1
Female	71.3 ± 29.2
Male	94.7 ± 33.2
BMI (kg/m^2)	31.4 ± 8.1
BMI percentile[a] (%)	87.1 ± 23.4
Etiology of liver disease	
NAFLD	28 (66.7)
Other	9 (21.4)
Fontan	5 (11.9)
PDFF (%)	
Breath-held	16.6 ± 13.1
Navigator-gated	17.0 ± 13.2
R2* (s^{-1})	
Breath-held	29.3 ± 4.7
Navigator-gated	29.6 ± 5.2

Values are reported as counts and percentages and means ± standard deviations

[a] $n = 34$

Table 3 Pearson correlation and Lin's concordance correlation coefficients for PDFF by ROI number by imaging acquisition technique

Technique	PDFF groups compared	Pearson correlation			Lin's correlation coefficient	
		r	95% CI	*p* value	r_c	95% CI
BH	1 ROI versus 4 ROI	0.9988	0.9977–0.9994	< 0.0001	0.9988	0.9978–0.9993
BH	2 ROI versus 4 ROI	0.9996	0.9993–0.9998	< 0.0001	0.9996	0.9993–0.9998
BH	3 ROI versus 4 ROI	0.9999	0.9998–0.9999	< 0.0001	0.9999	0.9998–0.9999
NG	1 ROI versus 4 ROI	0.9983	0.9969–0.9992	< 0.0001	0.9980	0.9965–0.9989
NG	2 ROI versus 4 ROI	0.9994	0.9990–0.9997	< 0.0001	0.9994	0.9989–0.9997
NG	3 ROI versus 4 ROI	0.9999	0.9998–0.9999	< 0.0001	0.9999	0.9998–0.9999

BH breath-held, *NG* navigator-gated, *CI* confidence interval

Regarding ROI size, there was no significant difference in mean ROI area between the breath-held and navigator-gated sequences ($P = 0.794$) [Fig. 3].

PDFF by breath-held versus navigator-gated techniques

There was very strong positive correlation between weighted mean PDFF measurements based on four ROIs obtained from the two sequences ($r = 0.997$, 95% CI 0.994–0.998, $p < 0.0001$ [Fig. 4]) with an almost perfect corresponding Lin's concordance correlation coefficient ($r_c = 0.997$, 95% CI 0.994–0.998). Despite this, there was a statistically significant mean difference in PDFF measurements between the two sequences (16.61 ± 13.28% for breath-held, 16.96 ± 13.18% for navigator-gated, $P = 0.0409$) with a mean bias of -0.34% between sequences by Bland–Altman analysis (SD, 1.1%, 95% limits of agreement [LOA], -2.4 to $+1.7$% [Fig. 5]). Correlation and Lin's concordance correlation remained very strong and almost perfect, respectively, between techniques when fewer ROIs were used for measurement of weighted mean PDFF [Table 4].

Subanalysis of patients with PDFF less than 9%

Eighteen of 42 patients (42.9%) had a measured PDFF less than 9% by both techniques; no patient had a PDFF less than 9% on one sequence but not the other. In this subgroup, there was only strong positive linear correlation between PDFF values based on the weighted mean of four ROIs ($r = 0.615$, 95% CI 0.208–0.841, $P = 0.0066$ [Fig. 6]). Lin's concordance correlation coefficient demonstrated poor absolute agreement, however, between the breath-held and navigator-gated sequences ($r_c = 0.543$, 95% CI 0.170–0.780), and there was a significant difference in measured PDFF between techniques (4.37 ± 1.18% for breath-held, 4.97 ± 1.26% for navigator-gated, $P = 0.0285$). Bland–Altman analysis showed a mean bias of -0.6% between sequences (SD, 1.1%, 95% LOA, -2.7 to $+1.5$%).

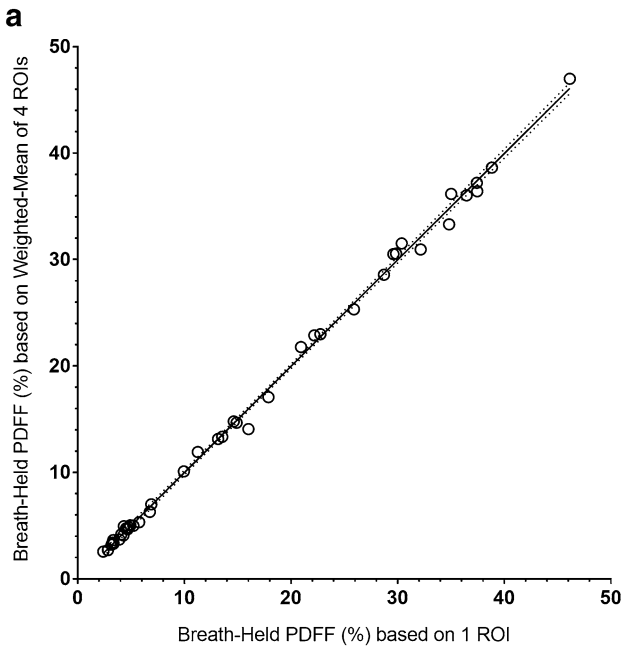
R2* by breath-held versus navigator-gated techniques

There was very strong positive correlation between weighted mean R2* measurements using four ROIs obtained by breath-held and navigator-gated techniques ($r = 0.837$, 95% CI 0.716–0.910, $P < 0.0001$) but with corresponding poor absolute agreement by Lin's concordance correlation coefficient ($r_c = 0.832$, 95% CI 0.712–0.905 [Fig. 7]). However, there was no statistically significant difference in weighted mean R2* measurements for the two techniques (29.3 ± 4.7 s^{-1} for breath-held, 29.6 ± 5.2 s^{-1} for navigator-gated, $P = 0.549$). Bland–Altman analysis revealed a mean bias of -0.27 s^{-1} between sequences (95% LOA, -6.0 to $+5.4$ s^{-1} [Fig. 8]).

Qualitative motion assessment

There was a significant difference in the frequency of clinically limiting motion between images acquired with the breath-held and navigator-gated techniques (47.7% [20/42] of breath-held image sets had clinically limiting motion versus 88.1% [37/42] of the navigator-gated image sets, $P = 0.0001$). Univariable and multivariable linear regression showed no association between age, sex, BMI, or weighted mean PDFF and the presence of qualitative clinically limiting motion on the breath-held sequence.

For 17 of the 42 patients (40.4%), qualitative motion scoring was categorically different (not clinically limiting versus clinically limiting) between the breath-held and navigator-gated sequences. In all of these cases, clinically limiting motion was present on the navigator-gated but not the breath-held sequence. Comparing the motion-concordant and discordant (between sequences) cohorts, there was no significant difference in the difference in measured weighted mean PDFF between sequences (difference of 0.66% in



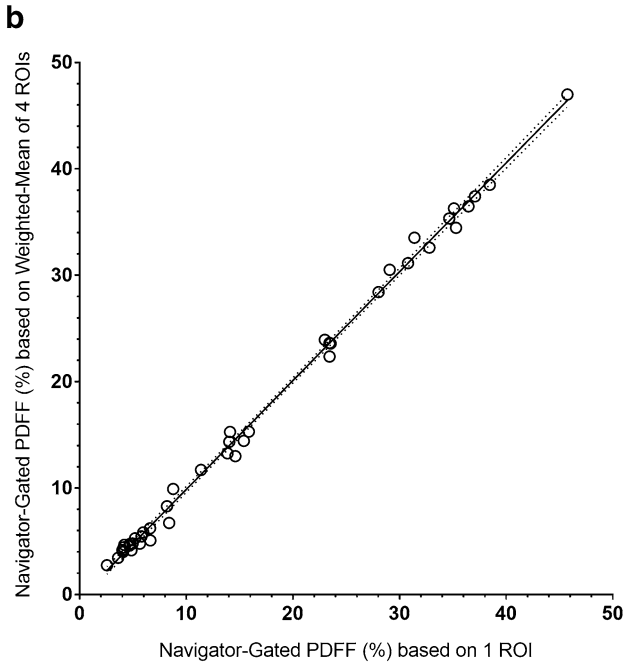


Fig. 2 Scatterplot showing weighted mean liver PDFF by four ROIs versus a single ROI for the breath-held (**a**) and navigator-gated sequences (**b**)

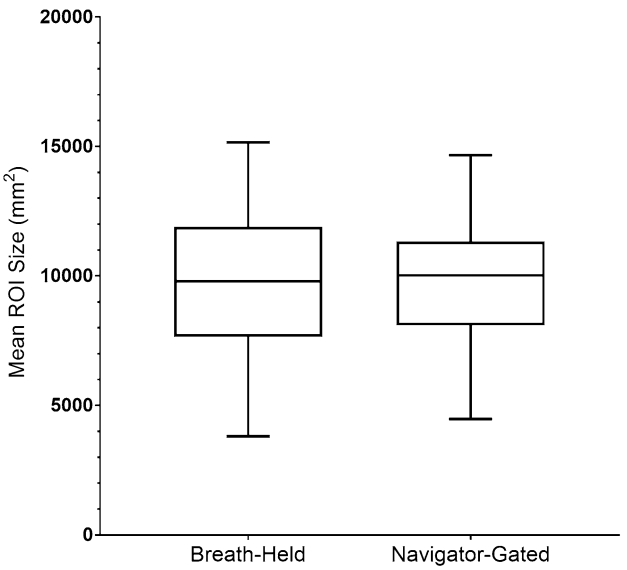


Fig. 3 Box and whisker plot of mean total ROI size for liver PDFF measurement by breath-held versus navigator-gated techniques

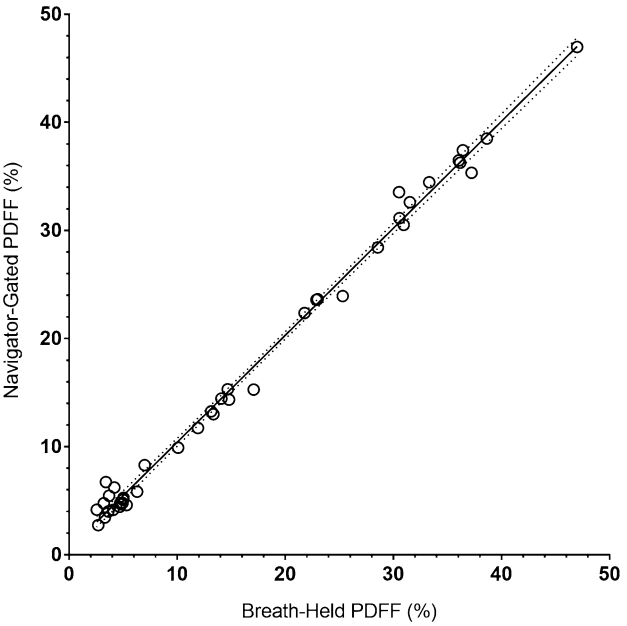


Fig. 4 Scatterplot of weighted mean liver PDFF by four ROIs for the breath-held versus navigator-gated techniques. 95% confidence interval demonstrated by dashed lines

the concordant cohort, 0.90% in the discordant cohort, $P = 0.3338$). That is to say, the discrepancy in measured PDFF between sequences was neither significantly greater nor smaller in patients who had substantial differences in motion between the two techniques.

Discussion

Multi-echo Dixon sequences like IDEALIQ® (GE Healthcare, Waukesha, WI), mDixon Quant® (Philips Healthcare, Best, The Netherlands), and q-Dixon® (Siemens Healthineers, Erlangen, Germany) can non-invasively estimate

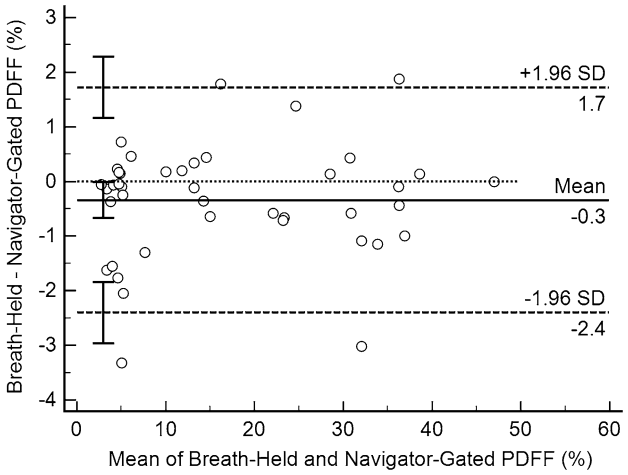


Fig. 5 Bland-Altman plot of weighted mean liver PDFF based on four ROIs for the breath-held versus navigator-gated techniques. Brackets indicate 95% confidence intervals

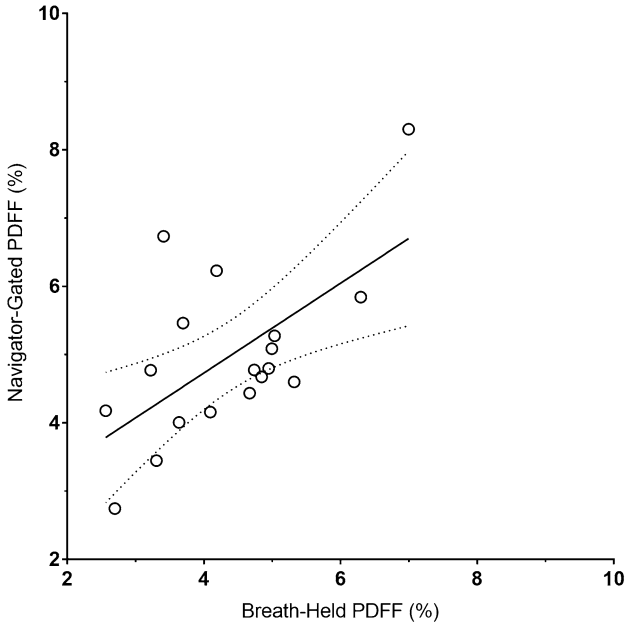


Fig. 6 Liver PDFF measured by a weighted mean of four ROIs for the breath-held versus navigator-gated techniques in the PDFF < 9% subgroup. 95% confidence interval demonstrated by dotted lines

liver fat and iron-related signal decay through measurement of PDFF and T2* (or R2*) values. These sequences are typically acquired in a single breath hold which, as implemented on the MRI scanners in this study, lasts approximately 25 s. Liver PDFF and T2* (or R2*) serve as imaging biomarkers and thereby impact the medical care of patients, including children, with liver disease [28, 29]. As such, it is important that accurate data are acquired even in patients with limited breath holding capacity. Respiratory triggering of an mDixon sequence could provide theoretical benefits of avoiding anesthesia and/or obtaining higher quality images in populations with limited breath holding capacity, such as children and some adults. To date, there are limited data regarding the use of respiratory compensation with navigator pulses to perform free breathing quantitative multi-echo Dixon imaging, particularly in children. One study in adults showed that respiratory gating techniques, including bellows and navigator echoes, provide valid PDFF and R2* values compared with breath-held images [30].

Our results demonstrate that, in a population of children and young adults with a range of hepatic fat fraction values (approximately 2–48%), qualitatively severe respiratory motion artifacts are more frequent when using a navigator-gated technique than a breath-held technique, despite the approximately 25 s breath hold required for the breath-held sequence. These results contradict the findings of a prior study of adult patients by Motosugi et al. which demonstrated no significant difference in image quality between mDixon sequences acquired with breath holding versus respiratory gating with bellows versus respiratory gating with a navigator echo [30]. We do note, however, that in that study, "poor image quality" was seen more often on images acquired with respiratory triggering, though the difference was not statistically significant [30]. The explanation for our finding of significantly increased frequency of motion artifact with respiratory triggering is uncertain, but further optimization of the sequence and acquisition is clearly warranted.

Table 4 Pearson correlation and Lin's correlation constant coefficients for PDFF by ROI number for breath-held versus navigator-gated techniques

PDFF groups compared	Pearson correlation			Lin's correlation coefficient	
	r	95% CI	*P*-value	r_c	95% CI
1 ROI BH versus 1 ROI NG	0.993	0.986–0.996	< 0.0001	0.992	0.985–0.996
2 ROI BH versus 2 ROI NG	0.995	0.990–0.997	< 0.0001	0.994	0.990–0.997
3 ROI BH versus 3 ROI NG	0.996	0.993–0.998	< 0.0001	0.996	0.992–0.998
4 ROI BH versus 4 ROI NG	0.997	0.994–0.998	< 0.0001	0.997	0.994–0.998

BH breath-held, *NG* navigator-gated, *CI* confidence interval

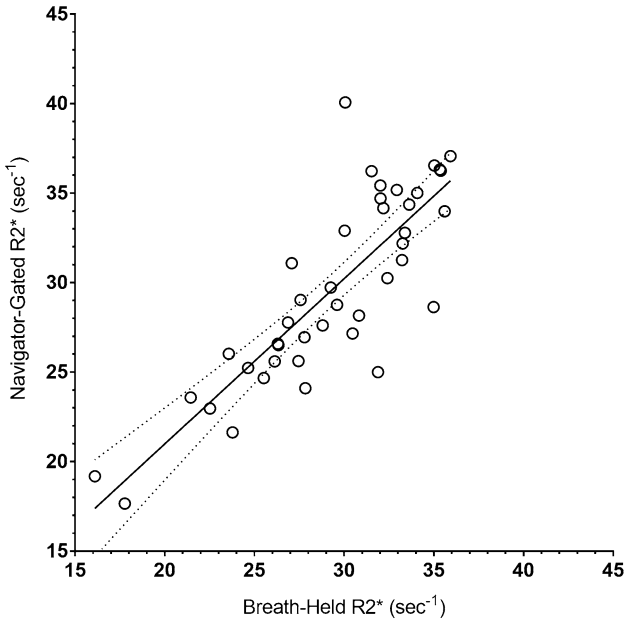


Fig. 7 Scatterplot of weighted mean liver R2* by four ROIs for the breath-held versus navigator-gated techniques. 95% confidence interval demonstrated by dotted lines

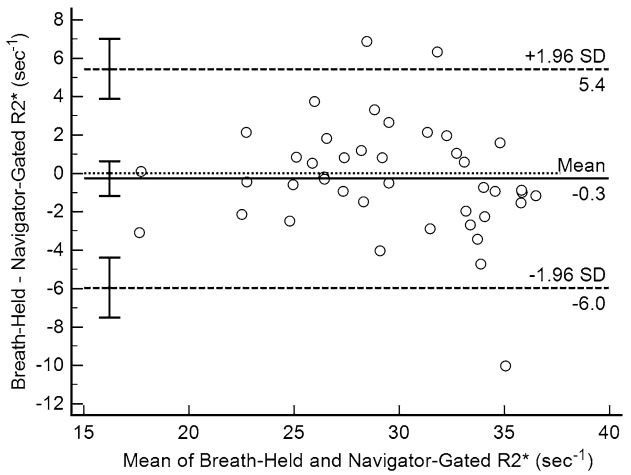


Fig. 8 Bland–Altman plot of weighted mean liver R2* based on four ROIs for the breath-held versus navigator-gated techniques. Brackets indicate 95% confidence intervals

Despite qualitatively worse respiratory motion artifact on the respiratory-compensated images, there was no significant difference in mean measureable ROI size between the breath-held and navigator-gated sequences, and there was very strong linear correlation and almost perfect Lin's concordance correlation ($r_c = 0.9965$) between weighted mean PDFF values derived from the two sequences for the population as a whole. Mean bias was only 0.3% fat fraction between the two techniques, suggesting that, if areas of artifact are avoided when placing ROIs, respiratory motion artifact does not clinically significantly impact measured liver PDFF. To put this bias in context, prior research has shown the mean bias of PDFF between field strength and imager manufacturers to range between 0.9 and 2.6% (95% LOA ± 2.8%), which is higher than the bias observed in our study [25, 31]. Moreover, Middleton et al. showed that in a population of children with NAFLD, a change in PDFF of + 5.5% or − 11% was 90% specific for a change in histological steatosis grade [16]. This degree of clinically relevant change in measured PDFF is much greater than the bias in measured PDFF between techniques in our study. As such, the small bias observed between acquisition techniques is likely clinically inconsequential.

Of note, in the subset of patients with lower PDFF values, where measurements of PDFF are known to be more variable, correlation was only strong (versus very strong in the overall population), and absolute agreement based on Lin's concordance correlation was poor (versus almost perfect in the overall population) [25]. Notably, 95% CIs of the correlation measures for this subpopulation did not overlap with those of the population as a whole, suggesting that correlation/concordance between techniques is significantly weaker in patients with lower (normal and mildly elevated) liver fat fractions. This result likely reflects the known higher variability in PDFF measurements at lower PDFF values more than it does differences between the sequences being tested in our study [25]. We do note, however, that even in this subgroup, mean bias in PDFF was only 0.6% between the breath-held and navigator-gated techniques, a difference that likely is not clinically relevant unless it results in a patient being erroneously categorized as having (or not having) an abnormal fat fraction based on a PDFF threshold.

Similar to PDFF measurements, there was very strong linear correlation between measured R2* values on breath-held versus navigator-gated sequences in the population overall. However, the correlation coefficient for R2* was lower than what was observed for PDFF, and absolute agreement assessed using Lin's concordance correlation was poor. This suggests that measured R2* is less equivalent between breath-held and navigator-gated sequences and likely more impacted by respiratory motion artifact. Nonetheless, we do note that while correlation was weaker for R2* than for PDFF, there was no significant difference in weighted mean R2* between the sequences, suggesting that this lack of perfect agreement may not be clinically relevant.

Although PDFF is already being used as an endpoint in clinical trials [4, 17, 32–34], there is not yet consensus on optimal ROI placement, size, or shape for PDFF measurement [24]. Various techniques of ROI placement have shown accurate measurement of hepatic fat compared to fat fractions obtained by histology or MR spectroscopy [1–14, 35], and without a head-to-head comparison to support a

superior technique, there remains variation in practice. In our study, we used free-hand ROIs encompassing as much liver as possible and compared PDFF values measured from a single ROI versus the weighted average measurement for up to four large ROIs, each placed at different slice locations within the mid-liver. Our results demonstrated very strong correlation and almost perfect Lin's concordance correlation between these measurements, suggesting that the placement of more than one large ROI may not provide any better measurement of PDFF, except in cases of dramatic heterogeneity in fat deposition (which was not observed in any patient in our population). These findings support the prior work by Campo et al. that showed the narrowest limits of agreement (LOA) between observers for PDFF were obtained by placing the largest ROI in each of the nine liver Couinaud segments [24]. While in that study, all ROIs were circular and thus, may not be directly comparable, multiple large circular ROIs end up encompassing the bulk of the liver, similar to what we have achieved with free-hand ROIs in our study.

There are several limitations to our study. First is the small number of patients, limited age range of patients, and limited spectrum of liver diseases in our study population, which may limit the applicability of our findings. Notably, the mean age of our study population was 15.7 years, reflecting a population that may be better at breath holding compared to younger or older, more debilitated patients. An additional limitation is that only one reviewer drew ROIs for PDFF and R2* measurements; therefore, interrater agreement was not assessed. Of note, in a prior study by Serai et al., interrater agreement was shown to be excellent using a similar technique of ROI placement [25]. Finally, our method of ROI placement avoided discrete areas of motion artifact in order to be pragmatic and mimic what would be done clinically; however, this likely inadequately reflects differences in PDFF and R2* values that may be caused by any motion differences between the two sequences being compared.

In conclusion, we have shown that despite greater motion artifacts, a navigator-gated multi-echo Dixon sequence targeted at liver fat fraction measurement produces PDFF values that are almost perfectly concordant with values from a breath-held sequence and thus may be a useful clinical tool if breath holding cannot be achieved. Further, we have shown that a single large ROI produces similar PDFF values to multiple ROIs, obviating the need to place more than one ROI to measure hepatic PDFF.

Disclosures Jonathan Dillman, MD, MSc and Andrew Trout, MD receive research Grants from Canon Medical Systems and Siemens Healthineers, travel support from Philips Healthcare, and in-kind support of research from General Electric Healthcare. The remaining authors have no disclosures to report.

References

1. Sharma P, Altbach M, Galons JP, Kalb B, Martin DR (2014) Measurement of liver fat fraction and iron with MRI and MR spectroscopy techniques. Diagn Interv Radiol 20 (1):17-26. https://doi.org/10.5152/dir.2013.13124
2. Idilman IS, Aniktar H, Idilman R, Kabacam G, Savas B, Elhan A, Celik A, Bahar K, Karcaaltincaba M (2013) Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology 267 (3):767-775. https://doi.org/10.1148/radiol.13121360
3. Kukuk GM, Hittatiya K, Sprinkart AM, Eggers H, Gieseke J, Block W, Moeller P, Willinek WA, Spengler U, Trebicka J, Fischer HP, Schild HH, Traber F (2015) Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. Eur Radiol 25 (10):2869-2879. https://doi.org/10.1007/s00330-015-3703-6
4. Noureddin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le TA, Bettencourt R, Changchien C, Brenner DA, Sirlin C, Loomba R (2013) Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology 58 (6):1930-1940. https://doi.org/10.1002/hep.26455
5. Tang A, Desai A, Hamilton G, Wolfson T, Gamst A, Lam J, Clark L, Hooker J, Chavez T, Ang BD, Middleton MS, Peterson M, Loomba R, Sirlin CB (2015) Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. Radiology 274 (2):416-425. https://doi.org/10.1148/radiol.14140754
6. Zand KA, Shah A, Heba E, Wolfson T, Hamilton G, Lam J, Chen J, Hooker JC, Gamst AC, Middleton MS, Schwimmer JB, Sirlin CB (2015) Accuracy of multiecho magnitude-based MRI (M-MRI) for estimation of hepatic proton density fat fraction (PDFF) in children. J Magn Reson Imaging 42 (5):1223-1232. https://doi.org/10.1002/jmri.24888
7. Lin SC, Heba E, Bettencourt R, Lin GY, Valasek MA, Lunde O, Hamilton G, Sirlin CB, Loomba R (2017) Assessment of treatment response in non-alcoholic steatohepatitis using advanced magnetic resonance imaging. Aliment Pharmacol Ther 45 (6):844-854. https://doi.org/10.1111/apt.13951
8. Anderson LJ, Holden S, Davis B, Prescott E, Charrier CC, Bunce NH, Firmin DN, Wonke B, Porter J, Walker JM, Pennell DJ (2001) Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload. Eur Heart J 22 (23):2171-2179
9. Dixon WT (1984) Simple proton spectroscopic imaging. Radiology 153 (1):189-194. https://doi.org/10.1148/radiology.153.1.6089263
10. Idilman IS, Tuzun A, Savas B, Elhan AH, Celik A, Idilman R, Karcaaltincaba M (2015) Quantification of liver, pancreas, kidney, and vertebral body MRI-PDFF in non-alcoholic fatty liver disease. Abdom Imaging 40 (6):1512-1519. https://doi.org/10.1007/s00261-015-0385-0
11. Serai SD, Fleck RJ, Quinn CT, Zhang B, Podberesky DJ (2015) Retrospective comparison of gradient recalled echo R2* and spin-echo R2 magnetic resonance analysis methods for estimating liver iron content in children and adolescents. Pediatr Radiol 45 (11):1629-1634. https://doi.org/10.1007/s00247-015-3378-9
12. Stark DD, Moseley ME, Bacon BR, Moss AA, Goldberg HI, Bass NM, James TL (1985) Magnetic resonance imaging and spectroscopy of hepatic iron overload. Radiology 154 (1):137-142. https://doi.org/10.1148/radiology.154.1.3964933
13. Verlhac S, Morel M, Bernaudin F, Bechet S, Jung C, Vasile M (2015) Liver iron overload assessment by MRI R2* relaxometry in

highly transfused pediatric patients: an agreement and reproducibility study. Diagn Interv Imaging 96 (3):259-264. https://doi.org/10.1016/j.diii.2014.11.021
14. Wood JC (2007) Magnetic resonance imaging measurement of iron overload. Curr Opin Hematol 14 (3):183-190. https://doi.org/10.1097/MOH.0b013e3280d2b76b
15. Kramer H, Pickhardt PJ, Kliewer MA, Hernando D, Chen GH, Zagzebski JA, Reeder SB (2017) Accuracy of Liver Fat Quantification With Advanced CT, MRI, and Ultrasound Techniques: Prospective Comparison With MR Spectroscopy. AJR Am J Roentgenol 208 (1):92-100. https://doi.org/10.2214/AJR.16.16565
16. Middleton MS, Van Natta ML, Heba ER, Alazraki A, Trout AT, Masand P, Brunt EM, Kleiner DE, Doo E, Tonascia J, Lavine JE, Shen W, Hamilton G, Schwimmer JB, Sirlin CB, Network NCR (2018) Diagnostic accuracy of magnetic resonance imaging hepatic proton density fat fraction in pediatric nonalcoholic fatty liver disease. Hepatology 67 (3):858-872. https://doi.org/10.1002/hep.29596
17. Permutt Z, Le TA, Peterson MR, Seki E, Brenner DA, Sirlin C, Loomba R (2012) Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease - MRI accurately quantifies hepatic steatosis in NAFLD. Aliment Pharmacol Ther 36 (1):22-29. https://doi.org/10.1111/j.1365-2036.2012.05121.x
18. Reeder SB, Robson PM, Yu H, Shimakawa A, Hines CD, McKenzie CA, Brittain JH (2009) Quantification of hepatic steatosis with MRI: the effects of accurate fat spectral modeling. J Magn Reson Imaging 29 (6):1332-1339. https://doi.org/10.1002/jmri.21751
19. Lee SS, Lee Y, Kim N, Kim SW, Byun JH, Park SH, Lee MG, Ha HK (2011) Hepatic fat quantification using chemical shift MR imaging and MR spectroscopy in the presence of hepatic iron deposition: validation in phantoms and in patients with chronic liver disease. J Magn Reson Imaging 33 (6):1390-1398. https://doi.org/10.1002/jmri.22583
20. Sofue K, Mileto A, Dale BM, Zhong X, Bashir MR (2015) Interexamination repeatability and spatial heterogeneity of liver iron and fat quantification using MRI-based multistep adaptive fitting algorithm. J Magn Reson Imaging 42 (5):1281-1290. https://doi.org/10.1002/jmri.24922
21. Towbin AJ, Serai SD, Podberesky DJ (2013) Magnetic resonance imaging of the pediatric liver: imaging of steatosis, iron deposition, and fibrosis. Magn Reson Imaging Clin N Am 21 (4):669-680. https://doi.org/10.1016/j.mric.2013.05.001
22. Jaimes C, Kirsch JE, Gee MS (2018) Fast, free-breathing and motion-minimized techniques for pediatric body magnetic resonance imaging. Pediatr Radiol 48 (9):1197-1208. https://doi.org/10.1007/s00247-018-4116-x
23. Wang Y, Rossman PJ, Grimm RC, Riederer SJ, Ehman RL (1996) Navigator-echo-based real-time respiratory gating and triggering for reduction of respiration effects in three-dimensional coronary MR angiography. Radiology 198 (1):55-60. https://doi.org/10.1148/radiology.198.1.8539406
24. Campo CA, Hernando D, Schubert T, Bookwalter CA, Pay AJV, Reeder SB (2017) Standardized Approach for ROI-Based Measurements of Proton Density Fat Fraction and R2* in the Liver. AJR Am J Roentgenol 209 (3):592-603. https://doi.org/10.2214/AJR.17.17812
25. Serai SD, Dillman JR, Trout AT (2017) Proton Density Fat Fraction Measurements at 1.5- and 3-T Hepatic MR Imaging: Same-Day Agreement among Readers and across Two Imager Manufacturers. Radiology 284 (1):244-254. https://doi.org/10.1148/radiol.2017161786
26. Evans JD (1996) Straightforward statistics for the behavioral sciences. Brooks/Cole Pub. Co., Pacific Grove
27. Lin LI (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics 45 (1):255-268
28. Anderson EL, Howe LD, Jones HE, Higgins JP, Lawlor DA, Fraser A (2015) The Prevalence of Non-Alcoholic Fatty Liver Disease in Children and Adolescents: A Systematic Review and Meta-Analysis. PLoS One 10 (10):e0140908. https://doi.org/10.1371/journal.pone.0140908
29. Schwimmer JB, Deutsch R, Kahen T, Lavine JE, Stanley C, Behling C (2006) Prevalence of fatty liver in children and adolescents. Pediatrics 118 (4):1388-1393. https://doi.org/10.1542/peds.2006-1212
30. Motosugi U, Hernando D, Bannas P, Holmes JH, Wang K, Shimakawa A, Iwadate Y, Taviani V, Rehm JL, Reeder SB (2015) Quantification of liver fat with respiratory-gated quantitative chemical shift encoded MRI. J Magn Reson Imaging 42 (5):1241-1248. https://doi.org/10.1002/jmri.24896
31. Kang GH, Cruite I, Shiehmorteza M, Wolfson T, Gamst AC, Hamilton G, Bydder M, Middleton MS, Sirlin CB (2011) Reproducibility of MRI-determined proton density fat fraction across two different MR scanner platforms. J Magn Reson Imaging 34 (4):928-934. https://doi.org/10.1002/jmri.22701
32. Jayakumar S, Middleton MS, Lawitz EJ, Mantry PS, Caldwell SH, Arnold H, Diehl AM, Ghalib R, Elkhashab M, Abdelmalek MF, Kowdley KV, Djedjos CS, Xu R, Han L, Subramanian GM, Myers RP, Goodman ZD, Afdhal NH, Charlton MR, Sirlin CB, Loomba R (2018) Longitudinal Correlations Between MRE, MRI-PDFF, and Liver Histology in Patients with Nonalcoholic Steatohepatitis: Analysis of Data from a Phase 2 Trial of Selonsertib. J Hepatol. https://doi.org/10.1016/j.jhep.2018.09.024
33. Loomba R, Sirlin CB, Ang B, Bettencourt R, Jain R, Salotti J, Soaft L, Hooker J, Kono Y, Bhatt A, Hernandez L, Nguyen P, Noureddin M, Haufe W, Hooker C, Yin M, Ehman R, Lin GY, Valasek MA, Brenner DA, Richards L, San Diego Integrated NRC (2015) Ezetimibe for the treatment of nonalcoholic steatohepatitis: assessment by novel magnetic resonance imaging and magnetic resonance elastography in a randomized trial (MOZART trial). Hepatology 61 (4):1239-1250. https://doi.org/10.1002/hep.27647
34. Middleton MS, Heba ER, Hooker CA, Bashir MR, Fowler KJ, Sandrasegaran K, Brunt EM, Kleiner DE, Doo E, Van Natta ML, Lavine JE, Neuschwander-Tetri BA, Sanyal A, Loomba R, Sirlin CB, Network NCR (2017) Agreement Between Magnetic Resonance Imaging Proton Density Fat Fraction Measurements and Pathologist-Assigned Steatosis Grades of Liver Biopsies From Adults With Nonalcoholic Steatohepatitis. Gastroenterology 153 (3):753-761. https://doi.org/10.1053/j.gastro.2017.06.005
35. Kim KY, Song JS, Kannengiesser S, Han YM (2015) Hepatic fat quantification using the proton density fat fraction (PDFF): utility of free-drawn-PDFF with a large coverage area. Radiol Med 120 (12):1083-1093. https://doi.org/10.1007/s11547-015-0545-x